Title: Communicative informativeness in aphasia: Investigating the relationship between linguistic and perceptual measures.

Authors: Janet Webster & Julie Morris

Addresses: Speech and Language Sciences, School of Education, Communication and Language Sciences, King George VI Building, Newcastle University, Newcastle upon Tyne, NE1 7RU

Short Title: Communicative informativeness in aphasia

Corresponding Author: Janet Webster

Contact Details: Speech and Language Sciences, School of Education, Communication and Language Sciences, King George VI Building, Newcastle University, Newcastle upon Tyne, NE1 7RU.

Telephone: 0191 208 5235 Fax: 0191 208 6518

Email: janet.webster@newcastle.ac.uk

Abstract

Purpose: Informativeness refers to how successfully a person is able to convey their intended message. This study explores the relationship between perceptual ratings of informativeness and selected linguistic measures of lexical and structural content. It considers which linguistic measures have ecological validity in terms of what listeners view as important.

Method: Two complex picture description samples from 20 people with aphasia were analysed. Linguistic measures included number of Correct Information Units (NCIU), percentage CIU (%CIU), number of propositions (NP), propositional idea density (PID) and mean length of utterance in words (MLU-w). Eleven naïve listeners produced direct magnitude estimation (DME) ratings of informativeness. A correlational design was used to investigate the relationship between mean DME informativeness ratings and each of the linguistic measures.

Results: The two picture description samples elicited similar informativeness ratings. Positive significant correlations were identified between mean DME informativeness ratings and NCIU, %CIU, NP and MLU-w; the strength of correlation differed across variables. No significant correlation was found between mean DME informativeness ratings and PID. Significant correlations were also seen between the linguistic variables, particularly between NP and PID and NP and MLU-w.

Conclusions: Overall, the linguistic measures corresponded to rated informativeness, highlighting their ecological validity. The strongest relationship was between NCIU and rated informativeness, emphasizing the importance of complete and accurate production of lexical information, particularly nouns. Less strong, but still significant, relationships were seen with variables looking at the efficiency of information giving and the connection of ideas within sentences. The importance of different types of informative measures is considered in relation to the elicitation stimuli.

Introduction

Within the assessment of people with aphasia, it is important to consider how successfully an individual is able to convey their intended message; an aspect referred to as informativeness (Oelschlaeger & Thorne, 1999). Measures that attempt to quantify informativeness 'may be the most valid indicators of the success with which speakers with aphasia communicate in daily life' (Doyle, Goda, & Spencer, 1995, p53). Over recent years, there has been an increased focus on measuring informativeness, with the emergence of studies investigating a range of linguistic measures and perceptual ratings of communication. There has been limited investigation of the relationship between these two types of measure.

Informativeness

Communicative success depends on the quantity and quality of information conveyed by a speaker and the efficiency of transmission (Nicholas & Brookshire, 1993). Informativeness has been considered in a range of discourse types: description of complex pictures, picture sequences, requests for personal and procedural information (Doyle, Tsironas, Goda, & Kalinyak, 1996; Nicholas & Brookshire, 1993), Cinderella narratives (Cupit, Rochon, Leonard, & Laird, 2007, 2010; Jacobs, 2001), interview samples (Bryant et al., 2013), elicited conversation in a simulated natural environment (Doyle et al., 1995) and natural conversation (Oelschlaeger & Thorne, 1999). Whilst measuring communicative informativeness and efficiency in conversation may be the most valid measure of the communication abilities of an individual with aphasia, conversation samples are variable and collecting and analysing them in a clinical setting is possibly impractical (Doyle et al., 1995). It can also be difficult to determine the accuracy and appropriacy of information in conversation, when listeners are making judgments based on the context of the utterance (Oelschlaeger & Thorne, 1999). Picture description samples are less naturalistic but provide more consistency regarding the potential target, allow greater control of variability and are efficient and clinically feasible (Doyle et al., 1995). Doyle et al., (1995) investigated the relationship between informativeness of structured elicitation procedures (as described by Nicholas & Brookshire, 1993) and elicited conversation. Correlational data revealed a strong positive association, suggesting that structured procedures (including picture description) can be used to predict informativeness during conversation.

Linguistic Measures

There are a variety of linguistic analyses that have considered language productivity, information content and grammatical complexity in the discourse of people with aphasia (see Bryant, Ferguson, & Spencer, 2016, for a review). This section will consider specific measures that have been utilised in relation to communicative informativeness and efficiency.

Analyses of Information Content

Informativeness, by definition, is related to the information the person with aphasia is able to convey successfully. Measures of information content have focused on the quantity of information as well as the efficiency with which information is conveyed. In considering informativeness, the most dominant analysis has been the identification of Correct Information Units (CIU) and the associated measures of communicative efficiency (percentage CIU in relation to overall number of words and CIU per minute) (Nicholas & Brookshire, 1993). The CIU analysis is rule-based, with rules to identify words and CIU; CIU are words that are informative, accurate and relevant in relation to the elicitation stimulus. The analysis can, therefore, be applied across different stimuli. The original study (Nicholas & Brookshire, 1993) applied the procedure to picture description (both single pictures and picture sequences) and personal and procedural recounts. The analysis was shown to be reliable (95% intra-rater reliability, 90% inter-rater reliability) and sensitive, as it distinguished between people with aphasia and normal speakers. As a group, speakers with no brain damage produced more words, more CIU, a higher percentage of CIU, more words per minute and more CIU per minute compared to people with aphasia. There was, however, some overlap in the scores between individuals in the two groups across all measures. Subsequent studies have shown that CIU related measures are related to aphasia severity (Ross & Wertz, 1999), can be applied to other elicitation conditions (e.g. elicited conversations, Doyle et al., 1995) and can be sensitive to changes seen post-intervention (e.g. Jacobs, 2001). Reliability has, however, varied across studies (Doyle et al., 1995; Jacobs, 2001) possibly due to the amount of training and discussion available prior to agreement. When applied to natural conversation, there was poor reliability (73% intra-rater, 55% inter-rater reliability), suggesting that the analysis is not currently appropriate for conversational discourse (Oelschlaeger & Thorne, 1999). The CIU

analysis contrasts with other analyses of information content which are stimulus specific, for example, content units (CU, Yorkston & Beukelman, 1980), accurate and complete main concepts (ACMC, Nicholas & Brookshire, 1995), main concept analysis (Dalton & Richardson, 2015; Richardson & Dalton, 2016). These measures consider the relative importance of information, as determined by the content of descriptions produced by speakers without aphasia.

Analyses of Propositional Content

CIU characterise information content but also represent a measure of relevant and accurate lexical content. Other ways of characterising lexical content may also be relevant when considering informativeness, including propositional content. Harley (2008) defines a proposition as 'the smallest unit of knowledge that can stand alone; it has a truth value – that is, a proposition can be either true or false' (p.379). Propositions include the verb and its arguments, adjectives, adverbs, prepositions and conjunctions (Brown, Snodgrass, Kemper, Herman, & Covington, 2008) and 'provide an index of semantic meaningfulness' (Bryant et al., 2013 , p993). Within discourse, the number of propositions provides a measure of the quantity of information being conveyed. Propositional idea density (PID), the number of propositions in relation to the total number of words, provides a measure of efficiency and reflects the ability to express relations between words (Fromm et al., 2016). PID is significantly reduced in the speech of people with aphasia compared to healthy speakers in procedural discourse and personal narratives (Fromm et al., 2016) and in interview samples (Bryant et al., 2013). Bryant et al., (2013) found a significant correlation between aphasia severity and PID, with a reduction in PID as severity of aphasia increased. Fromm et al. (2016) described a more complex relationship, with PID differentially sensitive to aphasia type and severity; a higher PID was associated both with increased severity (Wernicke's aphasia) and reduced severity (Anomic aphasia). Within both of these studies, propositional content was determined by an automated propositional analysis, Computerized Propositional Idea Density Rater (CPIDR, Brown et al., 2008). CPIDR can be used to provide an accurate and reliable analysis, which is significantly less time consuming but has good inter-rater reliability with manual calculation (Bryant et al., 2013). Bryant et al. (2013) considered PID alongside other discourse measures associated with informativeness as a way of validating the analysis. A reduction in PID, which distinguished

people with aphasia from healthy speakers, was accompanied by a reduced number of different words, reduced MLU and an increase in type token ratio. The relationship between PID and perceptual ratings of informativeness has not been considered.

Structural Analyses

Informativeness may also be related to the extent to which people with aphasia can produce sentences. There are many different measures of sentence production and complexity, for example, number of words in sentences, proportion of well-formed sentences (see Bryant et al., 2016). Mean length of utterance in words (MLU-w) is a simple but useful measure of verbal productivity and indirectly grammatical complexity (Dethorne, Johnson, & Loeb, 2005). When MLU is calculated in words, it is also strongly influenced by semantic content (Dethorne et al., 2005). The relationship between measures of sentence structure and informativeness has not been established.

Perceptual Measures

There are a number of studies that have used listener ratings to consider the features of and evaluate changes in the connected speech of people with aphasia. Within treatment studies (e.g. Cupit et al., 2007, 2010), ratings have been used as a means of social validation as they measure the social significance of changes in communicative ability and may be a 'proxy measure of the person's communicative success' (Cupit et al., 2010, p1488). Across studies, listeners have been asked to rate different elicitation samples, different features and use different types of rating scales. Cupit and colleagues (Cupit et al., 2007, 2010) asked listeners to rate pre- and post-therapy Cinderella narratives using seven point Likert scales for four discourse parameters (amount of information, ability to transmit the message, ability to find the words, degree of ease in retelling the narrative). There were strong correlations between the different scales so a single composite score was used to consider the impact of treatment. Hickey and Rondeau (2005) used 0 to 100% rating scales, with listeners asked to consider specific features (e.g. expression of information, turn taking, topic changes) and the overall quality of conversation. Other studies (e.g. Doyle et al., 1996; Jacobs, 2001) have used Direct Magnitude Estimation (DME) ratings to evaluate informativeness. DME is an experimental rating technique used to determine subjective estimates of the magnitude of a given variable

(Sorace, 2010). Doyle et al. (1996) suggested that DME is the preferable method for perceptual ratings as it has the necessary construct validity for detecting finely graded differences. Within DME, listeners assign a numerical value to an initial sample and then rate subsequent samples in relation to that; ratings from individuals are then converted to a common scale. In the Doyle et al. (1996) study, listeners rated overall informativeness. In the Jacobs (2001) study, listeners provided DME ratings for four constructs:- effectiveness, informativeness, grammaticality and listener comfort.

Within the studies, ratings have been obtained from a range of listeners, including speech and language pathologists (SLPs), speech and language pathology students, older and younger adults with no experience of aphasia. There is no consensus as to whether listener characteristics influence the ratings obtained, with mixed results about the effect of experience of aphasia/professional training and age/life experience. Hickey and Rondeau (2005) compared the ratings of SLPs, SLP students and naïve adults (mean age 42, range 18-72) for conversations between a person with aphasia and a student. Conversations were recorded before and after the students received training in supported communication. SLPs and SLP students provided higher ratings regarding quality of interaction compared to naïve listeners. The naïve listeners perceived the greatest amount of change post-training. Hickey and Rondeau (2005) conclude 'truly naïve judges who are representative of the general public may provide the most robust findings' (p39), suggesting that they focus on global communication. However, no graduated effect of professional training and experience was seen between SLPs and students, suggesting that other factors, for example, age and life experience may also be relevant. Cupit and colleagues (2010) compared SLPs, naïve younger adults and naïve older adults (of similar age to the SLPs). Although some interactions were seen between listener group and their ratings of change, the authors conclude that there were no important differences between listeners.

Relationship between Linguistic Measures and Perceptual Ratings of Informativeness

There have been a limited number of studies which have considered the relationship between linguistic measures and perceptual ratings of informativeness. Doyle et al. (1996) examined the relationship between listener judgments (DME ratings) and people with aphasia's production of CIU and Accurate and Complete Main Concepts (ACMC). 25 people with aphasia were included in the study, with a range of aphasia severity

and aphasia type (17 anomic, 4 Broca's, 2 conduction, 1 transcortical sensory and 1 transcortical motor) as defined by the Western Aphasia Battery (WAB, Kertesz, 2006). The elicitation stimuli described by Nicholas and Brookshire (1993) were used, with the calculation of number CIU, percentage CIU, CIU per minute and percentage ACMC. Eleven listeners then rated each description for overall informativeness. Listeners were older adults (60-78 years) with experience of working and communicating with chronically ill individuals. Correlations between the group mean DME for each participant, each of the linguistic measures and aphasia severity, as measured by the WAB Aphasia Quotient (WAB-AQ, Kertesz, 2006), were calculated. Significant positive correlations ($p<.01$) were identified between rated informativeness and WAB-AQ, number CIU, percentage CIU, CIU per minute and percentage ACMC. There were also significant correlations between each of the variables.

Ross and Wertz (1999) investigated the relationship between CIU and ACMC analyses and listener ratings of change. A picture description sample (WAB picnic scene, Kertesz, 2006) from 22 people with aphasia was obtained at two time points, two to nine months apart. Listeners judged whether the second sample they heard was better than, the same as or worse than the previous in terms of communicative ability. Changes in words per minute ($r=0.623$, $p<.05$), CIU per minute ($r=0.493$, $p<.05$) and percentage ACMC ($r=.543$, $p<.05$) were significantly related to listeners' judgment of change. There was no significant relationship between listeners' judgment and percentage CIU.

Jacobs (2001) looked at changes in informativeness pre- and post-therapy, with naïve listener ratings of Cinderella samples used to monitor the social validity of changes seen following sentence level therapy. Following treatment, statistically significant gains were seen in words per minute and CIU per min. These gains were accompanied by a corresponding (but not statistically significant) increase in listener ratings. Cupit and colleagues (2010) also considered the relationship between linguistic measures and listener ratings when investigating the social validity of treatment. They examined whether changes in listener ratings pre- and post-intervention corresponded to the presence of key propositions (events) within a narrative; propositions were identified from analysis of narratives from normal speakers. No statistical analysis was

carried out. However, within the groups of participants who showed change in listener ratings, a large proportion of participants produced a higher ratio of propositions. None of the participants in the control group (who showed no change in listener rating) showed change on the propositional analysis.

Summary

There has been some investigation of the relationship between linguistic measures and perceptual ratings of informativeness. It is, however, essential to consider the relationship with a broader range of linguistic measures to consider whether certain measures have stronger ecological validity in terms of what listeners view as important.

Aims

The study explores the relationship between linguistic measures and perceptual ratings of informativeness. The following research question was considered:

- How do listener ratings of informativeness using DME relate to a) Number of Correct Information Units (NCIU) b) Percentage Correct Information Units (%CIU) c) Number of propositions (NP) d) Propositional idea density (PID) and e) Mean length of utterance in words (MLU-w)?

The relationship was explored across two picture description elicitation contexts.

Methods

A correlational design was used to investigate bivariate correlations between mean informativeness DME ratings from naïve listeners and linguistic measures derived from picture description samples from people with aphasia. Some of this data was presented at the International Aphasia Rehabilitation Conference, Aveiro 2018, and is summarised in the conference proceedings (Webster, Harrison, & Morris, 2018).

Participants

Picture description samples of 20 people with aphasia were considered. Participants had been recruited to an intervention study (Morris, Howard, & Buerk, 2014) which investigated the impact of word retrieval therapies. Inclusion criteria for the original project were significant word retrieval difficulties (10-60% correct

on naming test), single symptomatic stroke, more than 3 months post-stroke, no significant wider cognitive difficulties and unimpaired or corrected hearing and vision that was adequate to participate in study. Participants could present with mild apraxia of speech or dysarthria if it was not the primary impairment. The study used data from the first 20 participants, including 17 men and 3 women, mean age 67.3 years (range 36 to 82 years) and mean time post onset 1.9 years (range 4 months to 10 years). Participants had aphasia of varying severity as indicated by the WAB AQ (Kertesz, 2006). Based on the WAB AQ criteria, there were 8 mild, 7 moderate, 4 severe and 1 person with very severe aphasia. From the WAB results, there were 7 people with Broca's aphasia (non-fluent), 7 Anomic (fluent), 5 Conduction (fluent) and 1 Wernicke's aphasia (fluent). Information about the individual participants can be found in table 1.

Insert table 1 here

Given the mixed findings regarding the impact of listener characteristics, naïve listeners were recruited as it was felt they would provide more objective ratings of broad communicative informativeness (Hickey & Rondeau, 2005). Listeners were two men and nine women who were current university students; they were aged between 19-65 and native speakers of English with little or no experience of aphasia.

Procedures

The anonymised samples were taken from the first, pre-therapy assessment phase in the original study (Morris et al., 2014) and consisted of two picture descriptions for each person with aphasia: the complex picture description from the Comprehensive Aphasia Test (CAT, Swinburn, Porter, & Howard, 2004) and the Tree picture (Nicholas & Brookshire, 1993). The complete samples were used as these were considered to be a closer approximation to natural communication (Cupit et al., 2007). Audacity® editing software was used to format and compile the audio data. Contributions from the assessor (e.g. instructions, questions, prompting and feedback) were removed. An audio file of the anonymous samples for each picture was produced. Within rating studies of complex, multi-dimensional variables (for example, intelligibility, Weismer & Laures, 2002), the choice of the initial sample has been shown to influence the DME scale value (although not the relative ordering of speakers), particularly if the standard is at the lower or upper end of the scale. In order

to compensate for any potential bias, an initial 'mid-range' sample (participant H) was selected to be the first sample. The researcher listened to all of the samples before any analysis was carried out; participant H subjectively seemed to be mid-range (i.e. not at the upper or lower end) in terms of the information conveyed. The other samples were then ordered randomly. A tone indicated the beginning and end of each sample, with eight seconds of silence between samples.

Listener Ratings

Listeners rated samples in a single session of around 120 minutes, including a 20 minute break. Sessions took place in a quiet room, with samples played over loud speakers. Prior to the ratings, participants were given a brief introduction to aphasia and informativeness and the task instructions. Samples for the CAT and Tree picture were rated in two blocks with the break between. Participants then had the opportunity to provide feedback about the task and to identify any factors they felt influenced their ratings. Ratings were obtained using the methodology described by Doyle et al (1996), with a rating sheet that included the picture stimulus for the samples being rated. Participants were asked to look at the picture and rate the informativeness of each sample, according to 'how accurately and completely the individual described the people, actions, setting and theme of the picture' (Doyle et al., 1996, p55). A DME rating procedure was used. Listeners were asked to rate the first sample, assigning any whole positive number above zero, and then rate each successive sample. They were told that low numbers indicated lower informativeness and higher numbers indicated higher informativeness.

Linguistic Analysis of Samples

The samples for the CAT picture and Tree picture were analysed separately. Samples were transcribed by one researcher using a standard method. Words and fillers were transcribed orthographically as complete words. Neologisms, phonological errors and word fragments were transcribed using broad phonemic transcription; if transcription was not possible, segments were coded as unintelligible. Transcriptions were then checked by a second researcher and the samples were then analysed considering the following measures.

Number of Correct Information Units (NCIU). The procedure described in Nicholas and Brookshire (1993) was used to identify words and CIU. CIU were words that were accurate, relevant and informative relative to the eliciting stimulus. Words did not have to be used in a grammatically accurate manner (Nicholas & Brookshire, 1993, p36).

Percentage Correct Information Units (%CIU). This measure was calculated using the formula: Number of CIU/Total number of words x 100.

Number of Propositions (NP). NP was calculated using CPIDR 5.1 (Covington, 2012). Previous studies have shown a high degree of reliability between CPIDR and manual counts of propositions (e.g. Bryant et al., 2013). CPIDR was used in 'speech mode' (as in Bryant et al., 2013), ensuring that repetitions were not counted as new propositions. Prior to the propositional analysis, the sample was cleaned (using the procedure described in Fromm et al., 2016), with the removal of utterances not directly related to the picture. In line with guidance about CIU analysis, phonological errors which were intelligible in context and common colloquialisms e.g. 'gonna' were replaced with the appropriate target words prior to analysis.

Propositional Idea Density (PID). This measure was also calculated using CPIDR 5.1 (Covington, 2012). It represented the number of expressed propositions divided by the number of words.

Mean Length of Utterance in Words (MLU-w) Samples were divided into utterances using the segmentation criteria (syntax, intonation, pauses, semantics) as outlined in Fromm et al. (2016). MLU-w was then calculated using the formula: Total number of words/Total number of utterances.

Reliability of Analysis

Ten percent (four transcripts) were randomly selected for intra-rater and inter-reliability ratings of CIU measures. Reliability of propositional measures was considered to be 100% due to the computerised nature of the analysis. For intra-rater reliability of CIU, the same researcher blind scored the samples at least two weeks after the original analysis. For inter-rater reliability, a second researcher blind scored the samples using the detailed instructions contained in Nicholas & Brookshire (1993). Point to point agreement was calculated (as in Doyle et al., 1996) as: Number of CIU agreements/Sum of CIU agreements and

disagreements x 100. The mean intra-rater reliability across the four transcripts was 91.3% (range 78.3-100%). The mean inter-rater reliability was 80.2% (range 62.1-89.3%). Discussion revealed that discrepancies arose due to the stricter application of semantic specificity and accuracy rules (2.11 & 3.11) and the repetition rule (2.14) by the initial rater. Inter-rater consensus was then reached and the agreed CIU analysis (as applied to all samples) fed into subsequent calculations. Within this study, the reliability of listener judgments was not considered; Doyle et al., (1996) showed that listener DME ratings made one week apart did not differ significantly from each other, showing that this type of rating is reliable.

Data Analysis

The raw listener ratings for each picture description were normalised onto one scale via geometric averaging (McGee, 2003); details of the procedure carried out can be found in Appendix 1. The mean normalised DME rating was then calculated for each sample for each participant with aphasia. The relationship between the mean DME rating and each linguistic variable for each sample was investigated using Spearman's rank correlations; due to the multiple comparisons for each dataset Bonferroni correction was used to consider statistical significance. In addition, correlations were carried out to consider the relationship between the linguistic variables.

Results

In order to understand if the two picture contexts elicited different DME informativeness ratings, the mean ratings were initially compared for the CAT and Tree picture; no significant difference was identified (Wilcoxon matched pairs, $z=1.47$, $p=0.140$, two tailed). The mean DME ratings and calculations for the linguistic variables can be found in Appendix 2.

Figure 1 and table 2 shows the relationship between the mean DME informativeness ratings and the linguistic variables for each picture. Significant correlations were found between the mean DME ratings and each of the linguistic variables with the exception of PID. Similar strengths of relationship were found across the CAT and Tree picture, with strength defined as in Evans (1996). A very strong relationship was found between NCIU and mean DME, with strong relationships for %CIU, NP and MLU-w.

Insert table 2 here

Insert figure 1 here

The relationship between the linguistic variables was then considered (see Table 3). Across both pictures, there was a significant positive correlation between NP and PID and NP and MLU-w. There were then correlations that were only significant for one of the pictures: PID and MLU-w (for Tree only), NCIU and %CIU, NCIU and NP and NCIU and MLU-w (for CAT only).

Insert table 3 here

Discussion

From this study, there is no evidence that the samples elicited via the two picture stimuli differed. There was no significant difference between the rated informativeness of the CAT and Tree picture and there were similar findings for the presence and strength of the relationships between the informativeness ratings and the linguistic measures across the pictures. Significant positive relationships were found between the DME informativeness ratings and the linguistic measures, with the exception of PID. This would suggest that NCIU, %CIU, NP and MLU-w all capture features which are related to listeners' perceptions of informativeness. There was a stronger relationship between NCIU compared to %CIU and between NCIU compared to NP. There was a significant correlation between the NP and MLU-w across both picture samples and a similarly strong association between these measures and rated informativeness. There were some relationships that were significant for only one picture. This may suggest some subtle differences between pictures but this should be interpreted with caution considering the small amount of data available and the number of comparisons carried out.

Quantity versus Efficiency of Information

This study replicates previous studies that have demonstrated the value of CIU analyses in the characterisation of informativeness. CIU measures (NCIU and %CIU) were both significantly related to the DME ratings of informativeness. There was, however, a stronger relationship with the quantity of

information (NCIU) compared to information efficiency (as captured by %CIU). This could be influenced by the instructions given to the listeners and the elicitation stimulus. They were asked to consider how 'accurately and completely…' people described the picture. With the picture present, the listeners had a clear sense of what was expected and were probably looking for detailed information about the scene. In this context, the relative prominence of information and the efficiency of production may have been considered less important. In other conditions, the balance between the value of information content and efficiency may be different. For example in conversation, the interaction between interlocutors requires more efficient exchange of information and %CIUs may be a more important measure of informativeness in this setting. The different combination of elicitation conditions in the Doyle et al. (1996) study may account for the more comparable strength of correlation they identified between informativeness and NCIU, %CIU and CIU per minute. It should, however, be noted that it was still NCIU that had the strongest correlation and which accounted for 33% of the variance seen within hierarchical multiple regression analyses. There needs to be further investigation of the potential value of CIU quantity and efficiency measures across different elicitation conditions.

Information Content versus Propositional Content

This study is the first to consider the relationship between propositional content and rated informativeness. A significant correlation was found between the number of propositions and informativeness. No significant association was seen between PID and rated informativeness. This mirrors the pattern discussed above, in terms of quantity versus efficiency of information. A stronger correlation was found between mean DME informativeness and NCIU compared to NP. CIU capture lexical information across word class but only words that are relevant and accurate. Within the picture descriptions, many of the CIU were nouns. In contrast, the propositional analysis only considered nouns when they were combined with a verb and included both accurate and relevant information and possibly inappropriate information. Either of these differences could account for the reduced strength of the relationship between NP and rated informativeness compared to NCIU.

Fromm and colleagues (2016) consider the complex relationship between propositional density, aphasia severity and type, with the critical factor being which words are included in the proposition count. The same factor needs to be considered when investigating the relationship between propositional measures and informativeness. Prior to the propositional analysis, responses that were not relevant to the task, for example, commentary on performance or questions about the task were removed. However, the speakers with aphasia could produce propositions that were inaccurate, for example, inappropriate prepositions, semantic errors, and these were still included in the analysis. In people with Wernicke's aphasia 'a higher PD score does not necessarily mean that discourse is more successful or less impaired' (Fromm et al., 2016, p1130). The less strong relationship between NP and rated informativeness (compared to CIU) could reflect the inclusion of inaccurate or irrelevant information. Two participants (participant B and S) produced more propositions than CIU (discrepancy of >10) across both samples suggesting they were producing inaccurate or irrelevant propositions. Interestingly, participant K who was classified as having Wernicke's aphasia did not seem to be producing a lot of inaccurate information. The number of propositions and propositional density may also be reduced in Broca's aphasia, as 'many grammatical elements that contribute to PD may be reduced or absent, whereas nouns which do not count as propositions, are more frequent' (Fromm et al., 2016, p1129). Nouns may be accurate, relevant and informative but will not by themselves count as propositions. The non-fluent participants with Broca's aphasia (with the exception of participant S) produced a very reduced number of propositions but there was no marked discrepancy between the number of CIU and propositions; they just produced very limited content overall. Four fluent participants (participants N, P, R & T) produced more CIU than propositions (discrepancy of >10) across both samples. This suggests that these participants were producing accurate and relevant information but not necessarily within a propositional or sentence context. The stronger relationship between rated informativeness and NCIU would suggest listeners were crediting the production of informative single words.

Sentence Structure and Propositional Content

This study is also the first to consider the relationship between MLU-w and rated informativeness, with a significant positive correlation identified. There was also a significant correlation between the number of

propositions and MLU-w across both picture samples and a similarly strong association between both of these measures and rated informativeness. It could be, therefore, that these measures are capturing very similar features of production, with consideration of not only the lexical content but the structure of the sentence. Both NP and MLU-w to some extent reflect the form of the message, considering the relationship between words. By definition, the non-fluent participants had very low MLU-W. The participants produced very restricted lexical content and words were generally produced in isolation or short phrases; there was very limited evidence of sentence structure.

Nicholas and Brookshire (1995) suggested that communicative success is more likely to depend on the informativeness of a speaker's message and the efficiency of conveying that message, than on the form. Research has shown that syntax training resulting in increased MLU does not necessarily result in changes in the amount of information conveyed as measured by CIU (Doyle & Bourgeois, 1986). Similarly, there is no straightforward relationship between linguistic measures of sentence structure and listener ratings of either grammaticality or overall communication (Jacobs, 2001). These studies conclude that following treatment, speakers who produce longer, more complex sentences may not become more adequate communicators. The findings of this study suggest a more complex picture, with a significant positive relationship between MLU-w and rated informativeness but a relationship that is not as strong as that with NCIU. As highlighted in the introduction, MLU-w is a measure of verbal productivity and grammatical complexity but also semantic content (Dethorne et al., 2005). The association between MLU-w and measures of informativeness is consistent with the recognised dependence of this measure on semantic content (Dethorne et al., 2005). However, similar to the propositional analysis, some of that semantic content may be appropriate and relevant to the picture, some may be inaccurate.

Final Considerations

This study investigated informativeness within a picture description task. Picture description is less naturalistic than conversation but provides consistency, a predictable target and is clinically feasible. As it provides a context with known targets, it may place an emphasis on the quantity of accurate and relevant information. In other contexts, listeners may have a different focus and it is important to investigate the

relationship between this range of measures and informativeness across a range of elicitation contexts, particularly narrative and elicited conversation. However, it is acknowledged that currently it may be problematic to consider naturalistic conversation due to the reliability of the CIU measures (Oelschlaeger & Thorne, 1999).

Within this study, the CIU and MLU-w analyses were carried out by hand. For CIU particularly, this was a time consuming process which also resulted in lower intra-rater and inter-rater reliability scores than in the original study (Nicholas & Brookshire, 1993). Reliability scores were, however, more consistent with other studies where the researchers have not received extensive initial training or had opportunity to discuss the application of the rules prior to analysis (Doyle et al., 1995; Oelschlaeger & Thorne, 1999). It would have been helpful to complete a final reliability score following the discussion of the discrepancies between raters. In both the original and current studies, it was subjectively harder to analyse the fluent speakers and it was a fluent speaker who accounted for the lowest scores for both intra- and inter-rater reliability. Whilst Doyle et al., (1996) reported reliability, nevertheless in future studies, it may also be important to consider the reliability of the perceptual judgements. Computerised analyses are more reliable and are also significantly less time consuming; this means that the propositional analysis via CPIDR (Covington, 2012) and automated analyses of MLU-w may be clinically more feasible. It should, however, be acknowledged that these analyses had a less strong relationship with rated informativeness.

Conclusion

This study has investigated the relationship between a range of linguistic measures and perceptual ratings of informativeness. It contributes to our knowledge about what factors influence a listener's perception about whether a person with aphasia has successfully communicated their message. NCIU, %CIU, NP and MLU-w all captured features which related to listeners' ratings of informativeness. They are, therefore, all ecologically valid measures which can be used to consider how effectively a person is able to convey information; which measure is chosen may depend on context and aspects of particular interest.

Acknowledgements

Thanks to Sarah Harrison who collected and analysed the data for this study as part of her undergraduate dissertation project. Thanks also to the participants in the original intervention study, to the listeners who were involved in this study and to Hannah Deards who carried out the inter-rater reliability.

Disclosure

No potential conflict of interest was declared by the authors.

Funding

This project used anonymised data originally collected during a word retrieval intervention project (SemaFore, Morris, Howard and Buerk 2014); this project was funded by the National Institute for Health Research's Research for Patient Benefit funding stream.

References

Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., & Covington, M. A. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior research methods, 40*(2), 540-545. doi:10.3758/BRM.40.2.540

Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical linguistics & phonetics*, 1-30. doi:10.3109/02699206.2016.1145740

Bryant, L., Spencer, E., Ferguson, A., Craig, H., Colyvas, K., & Worrall, L. (2013). Propositional Idea Density in aphasic discourse. *Aphasiology, 27*(8), 992-1009. doi:10.1080/02687038.2013.803514

Covington, M. A. (2012). CPIDR 5.1: Computerized Propositional Idea Density Rater. Georgia, USA: Institute for Artifical Intelligence, The University of Georgia.

Cupit, J., Rochon, E., Leonard, C., & Laird, L. (2007). Social validation: Examining its sensitivity and the factors that influence raters' judgments. *Brain and Language, 103*(1), 244-245. doi: 10.1016/j.bandl.2007.07.026

Cupit, J., Rochon, E., Leonard, C., & Laird, L. (2010). Social validation as a measure of improvement after aphasia treatment: Its usefulness and influencing factors. *Aphasiology, 24*(11), 1486-1500. doi:10.1080/02687031003615235

Dalton, S. G., & Richardson, J. D. (2015). Core-lexicon and main-concept production during picture-sequence description in adults without brain damage and adults with aphasia. *American Journal of Speech-Language Pathology, 24*(4), S923-S938. doi:10.1044/2015_AJSLP-14-0161

Dethorne, L. S., Johnson, B. W., & Loeb, J. W. (2005). A closer look at MLU: What does it really measure? *Clinical linguistics & phonetics, 19*(8), 635-648. doi:10.1080/02699200410001716165

Doyle, P. J., & Bourgeois, M. S. (1986). The effects of syntax training on "adequacy" of communication in Broca's aphasia: A social validation study. In R. H. Brookshire (Ed.), *Clinical Aphasiology Conference Proceedings* (pp. 123-131). Minneapolis: BRK.

Doyle, P. J., Goda, A. J., & Spencer, K. A. (1995). The communicative informativeness and efficiency of connected discourse by adults with aphasia under structured and conversational sampling conditions. *American Journal of Speech-Language Pathology, 4*(4), 130-134. doi:10.1044/1058-0360.0404.130

Doyle, P. J., Tsironas, D., Goda, A. J., & Kalinyak, M. (1996). The relationship between objective measures and listeners' judgments of the communicative informativeness of the connected discourse of adults with aphasia. *American Journal of Speech-Language Pathology, 5*(3), 53-60.

Evans, J. D. (1996). *Straightforward Statistics for the Behavioural Sciences*. Pacific Grove, CA: Brooks/Cole Publishing

Fromm, D., Greenhouse, J., Hou, K., Russell, G. A., Cai, X., Forbes, M., . . . MacWhinney, B. (2016). Automated Proposition Density Analysis for Discourse in Aphasia. *Journal of Speech, Language, and Hearing Research, 59*(5), 1123-1132. doi:10.1044/2016_JSLHR-L-15-0401

Harley, T. A. (2008). *The Psychology of Language: From Data to Theory* (Third ed.). Hove, East Sussex: Psychology Press.

Hickey, E., & Rondeau, G. (2005). Social validation in aphasiology: Does judges' knowledge of aphasiology matter? *Aphasiology, 19*(3-5), 389-398. doi:10.1080/02687030444000831

Jacobs, B. J. (2001). Social validity of changes in informativeness and efficiency of aphasic discourse following linguistic specific treatment (LST). *Brain and Language, 78*(1), 115-127. doi:10.1006/brln.2001.2452

Kertesz, A. (2006). *Western Aphasia Battery-Revised* San Antonio, Texas: Pearson.

McGee, M. (2003). *Usability magnitude estimation.* Paper presented at the Human Factors and Ergonomics Soceity Annual Meeting

Morris, J., Howard, D., & Buerk, F. (2014). *SemaFoRe: Comparing word retrieval treatments for aphasia via a randomised crossover trial.* Paper presented at the Clinical Aphasiology Conference, St Simon's Island, GA.

Nicholas, L. E., & Brookshire, R. H. (1993). A system of quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research, 36*, 338-350.

Nicholas, L. E., & Brookshire, R. H. (1995). Presence, completeness and accuracy of main concepts in the connected speech of non-brain damaged adults and adults with aphasia. *Journal of Speech and Hearing Research, 38*, 145-156. doi:10.1044/jshr.3801.145

Oelschlaeger, M. L., & Thorne, J. C. (1999). Application of the correct information unit analysis to the naturally occurring conversation of a person with aphasia. *Journal of Speech, Language, and Hearing Research, 42*(3), 636-648. doi:10.1044/jslhr.4203.636

Richardson, J. D., & Dalton, S. G. (2016). Main concepts for three different discourse tasks in a large non-clinical sample. *Aphasiology, 30*(1), 45-73. doi:10.1080/02687038.2015.1057891

Ross, K. B., & Wertz, R. T. (1999). Comparison of impairment and disability measures for assessing severity of, and improvement in, aphasia. *Aphasiology, 13*(2), 113-124. doi:10.1080/026870399402235

Sorace, A. (2010). Using magnitude estimation in language acquisition research. In S. Unswoth & E. Blom (Eds.), *Experimental Methods in Language Acquisition* Amsterdam: John Benjamins.

Swinburn, K., Porter, G., & Howard, D. (2004). *The Comprehensive Aphasia Test*. Hove: Psychology Press.

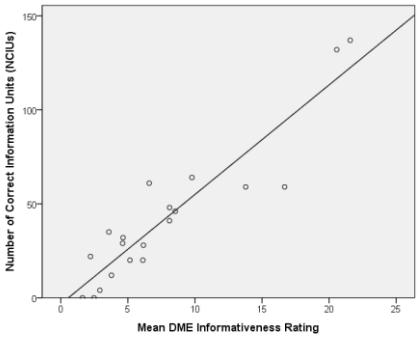
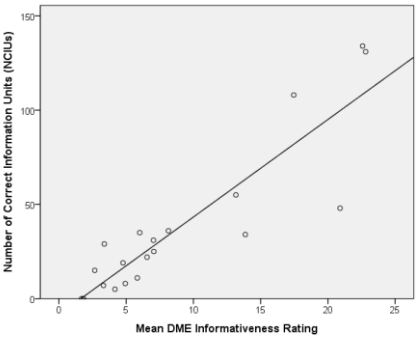
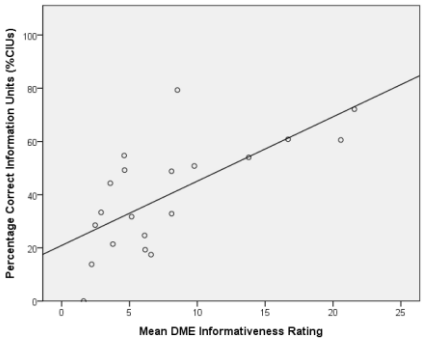
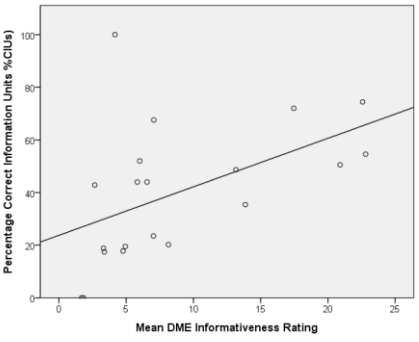
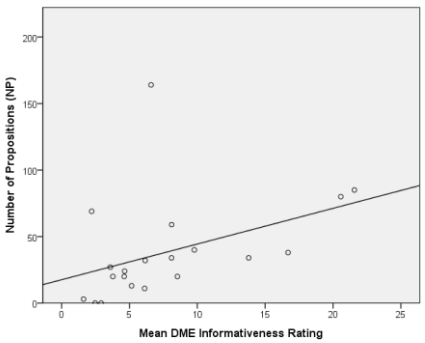
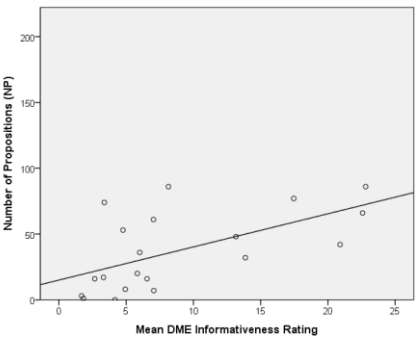
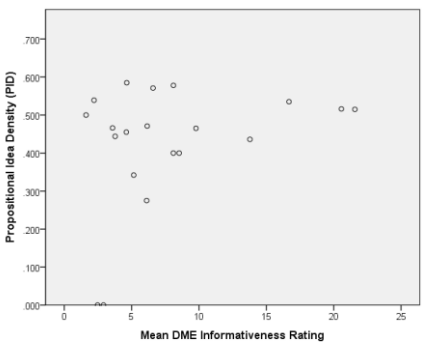
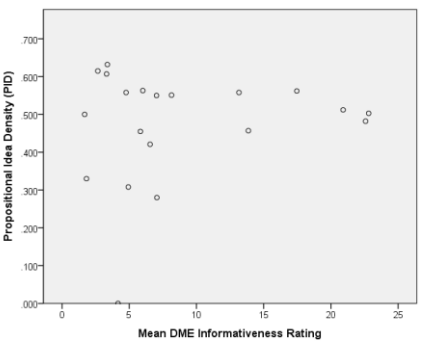
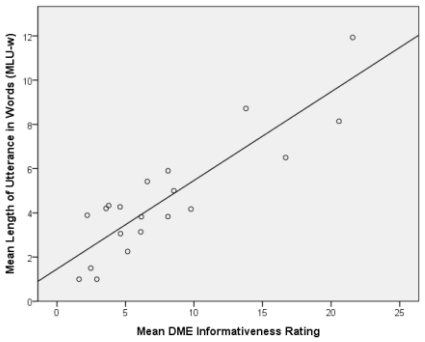
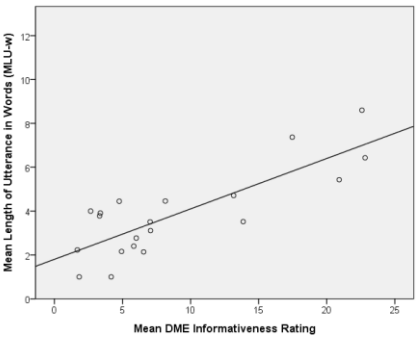
Webster, J., Harrison, S., & Morris, J. (2018). Communicative informativeness in aphasia. *Aphasiology, 32*(sup1), 243-245. doi:10.1080/02687038.2018.1485840

Weismer, G., & Laures, J. S. (2002). Direct magnitude estimates of speech inteligibility in dysarthira: Effects of a chosen standard. *Journal of Speech, Language and Hearing Research, 45*, 421-433. doi:10.1044/1092-4388(2002/033)

Yorkston, K. M., & Beukelman, D. R. (1980). An analysis of connected speech samples of aphasic and normal speakers. *Journal of speech and hearing disorders, 45*(1), 27-36. doi:10.1044/jshd.4501.27

Figure Legends

Figure 1: Relationship between mean DME informativeness ratings and the linguistic variables for the CAT and Tree pictures

	CAT Picture	Tree Picture
Number of Correct Information Units (NCIUs)		
Percentage Correct Information Units (%CIUs)		
Number of Propositions (NP)		
Propositional Idea Density (PID)		
Mean Length of Utterance in Words (MLU-w)		

Tables

Table 1: Participants with aphasia

Subject	**Age** (years)	**Time post-onset** (months)	**Gender**	**Handed-ness**	**WAB AQ**	**WAB Classification**	**WAB Fluency Score**	**Fluent/Non-Fluent**
A	64	36	Male	Right	70 moderate	Conduction	5	Fluent
B	70	46	Male	Right	84 mild	Conduction	8	Fluent
C	68	20	Male	Right	63 moderate	Broca's	2	Non-fluent
D	71	32	Male	Right	66 moderate	Conduction	6	Fluent
E	52	6	Male	Right	23 very severe	Broca's	0	Non-fluent
F	73	24	Male	Right	37 severe	Broca's	2	Non-fluent
G	36	34	Female	Right	76 mild	Anomic	8	Fluent
H	64	17	Male	Right	75 moderate	Conduction	9	Fluent
I	80	6	Female	Right	44 severe	Broca's	2	Non-fluent

J	58	18	Male	Right	28 severe	Broca's	1	Non-fluent
K	61	4	Male	Left	60 moderate	Wernicke's	5	Fluent
L	81	5	Male	Right	76 mild	Anomic	5	Fluent
M	78	9	Male	Right	69 moderate	Anomic	6	Fluent
N	82	5	Male	Right	83 mild	Anomic	9	Fluent
O	74	67	Male	Right	73 moderate	Broca's	4	Non-fluent
P	65	51	Female	Right	76 mild	Conduction	9	Fluent
Q	51	11	Male	Right	82 mild	Anomic	9	Fluent
R	75	120	Male	Right	80 mild	Anomic	6	Fluent
S	81	14	Male	Right	31 severe	Broca's	2	Non-fluent
T	67	5	Male	Right	90 mild	Anomic	9	Fluent

Table 2: Summary of correlations between mean DME informativeness ratings and the linguistic variables for the CAT and Tree pictures

	CAT Picture		Tree Picture	
	r value (strength)	**p value (significance)**	**r value (strength)**	**p value (significance)**
NCIU	.879 (very strong)	<.000*	.902 (very strong)	<.000*
%CIU	.695 (strong)	.001*	.652 (strong)	.002*
NP	.632 (strong)	.003*	.634 (strong)	.003*
PID	.217 (weak)	.357	-.068 (weak)	.777
MLU-w	.765 (strong)	<.000*	.651 (strong)	.002*

* Significant at Bonferroni corrected levels for multiple comparisons for each picture: p<.01. Strength of correlation defined as in Evans (1996).

Table 3: Summary of correlations between the linguistic variables for the CAT and Tree pictures

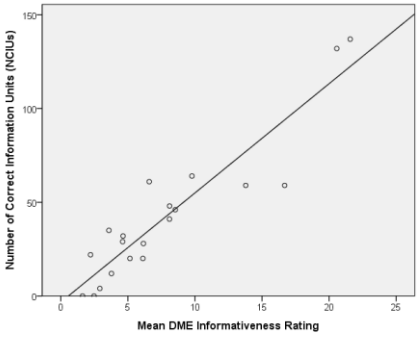
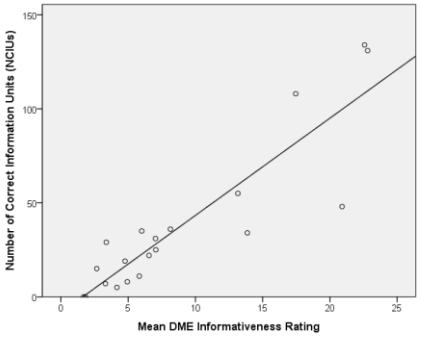
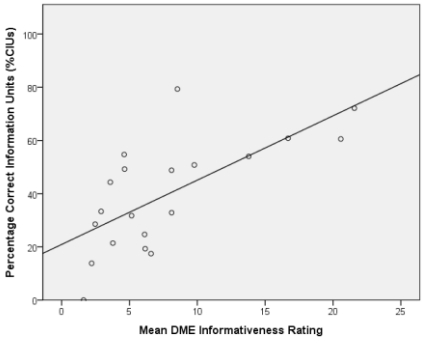
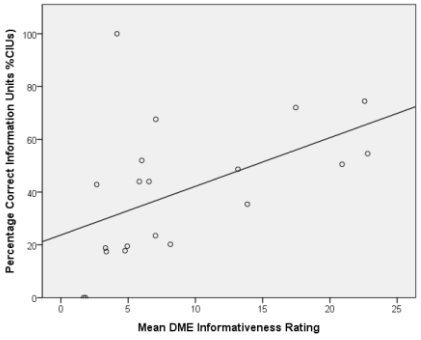
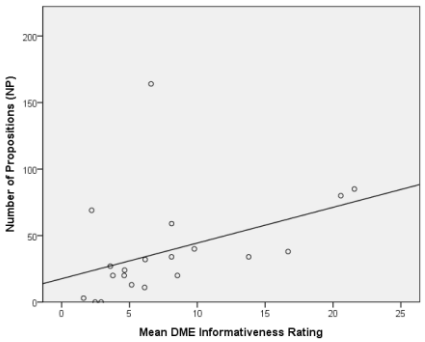
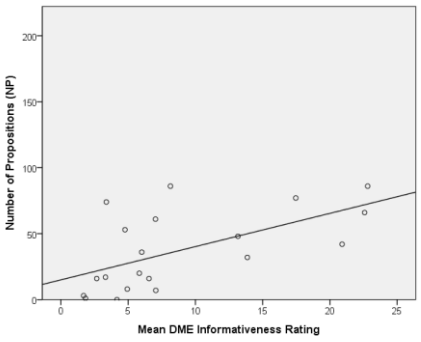
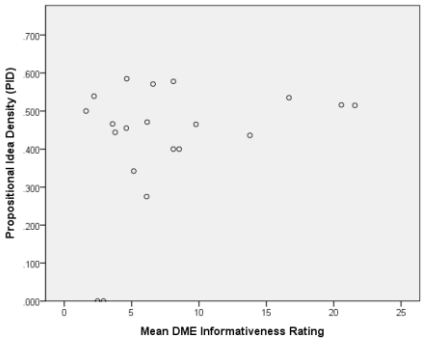
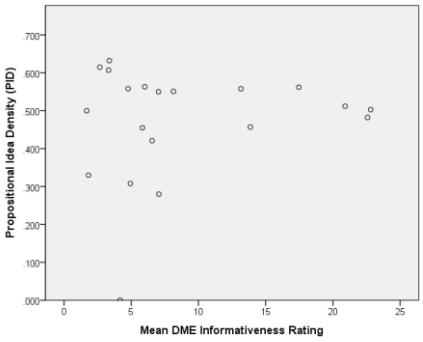
	CAT Picture				Tree Picture			
	%CIU	**NP**	**PID**	**MLU-w**	**%CIU**	**NP**	**PID**	**MLU-w**
NCIU	r=.636 p=.003*	r=.847 p<.001*	r=.482 p=.031	r=.826 p<.001*	r=.213 p=.367	r=.306 p=.190	r=.544 p=.011	r=.149 p=.529
%CIU		r=.226 p=.338	r=-.022 p=.927	r=.559 p=.010		r=.556 p=.011	r=.305 p=.190	r=.562 p=.010
NP			r=.708 p<.001*	r=.744 p<.001*			r=.813 p<.001*	r=.816 p<.001*
PID				r=.414 p=.069				r=.818 p<.001*

* Significant at Bonferroni corrected levels for multiple comparisons for each picture: p<.005.

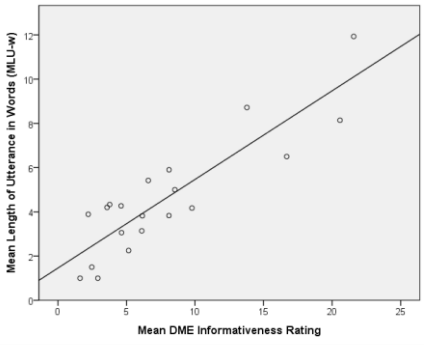
Appendix 1: Details of the process carried out to normalise informativeness ratings

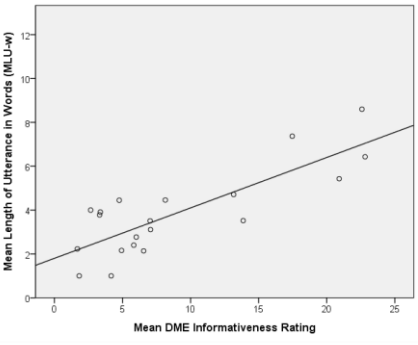
Details of the normalisation of data via geometric averaging for each picture; this followed the process described in McGee (2003).

1. Specification of raw DME rating for each listener for each participant with aphasia
2. Calculation of the log of each raw DME rating and then the mean log score for each listener
3. Calculation of overall mean log score (across listeners)
4. Subtraction of each mean listener's score from overall mean to calculate each listener's offset score
5. Standardisation of ratings by adding each listener's offset to each individual log score
6. Calculation of the antilog of the normalised log score, providing the normalised DME rating

	CAT Picture	Tree Picture
Number of Correct Information Units (NCIUs)		
Percentage Correct Information Units (%CIUs)		
Number of Propositions (NP)		
Propositional Idea Density (PID)		

Mean Length of Utterance in Words (MLU-w)





Appendix 2: Mean normalised DME informativeness ratings and linguistic variables for each picture

	CAT Picture						Tree Picture					
Subject	**Mean normalised DME rating**	**NCIUs**	**%CIUs**	**NP**	**PID**	**MLU-w**	**Mean normalised DME rating**	**NCIUs**	**%CIUs**	**NP**	**PID**	**MLU-w**
A	3.59	35	44.30	27	0.47	4.20	6.01	35	52.00	36	0.56	2.77
B	6.59	61	17.47	164	0.57	5.42	8.14	36	22.22	86	0.51	4.46
C	6.12	20	24.69	11	0.28	3.14	4.93	8	19.51	8	0.31	2.16
D	8.10	41	48.80	34	0.40	3.83	13.86	34	35.42	32	0.46	3.52
E	1.62	0	0	3	0.50	1.00	1.68	0	0.00	3	0.50	2.23
F	2.91	4	33.33	0	0	1.00	4.17	5	100.00	0	0.00	1.00
G	6.16	28	19.31	32	0.47	3.82	13.16	55	48.67	48	0.56	4.71
H	13.79	59	54.00	34	0.44	8.72	20.91	48	50.53	42	0.51	5.43
I	5.16	20	31.74	13	0.34	2.25	6.55	22	44.00	16	0.42	2.14

J	2.47	0	28.57	0	0	1.50	1.82	0	0.00	1	0.33	1.00
K	4.61	29	54.71	20	0.46	4.27	5.83	11	44.00	20	0.46	2.40
L	4.64	32	49.23	24	0.59	3.06	2.66	15	42.86	16	0.62	4.00
M	9.78	64	50.79	40	0.47	4.17	4.76	19	17.76	53	0.56	4.45
N	16.69	59	60.82	38	0.54	6.50	17.47	108	72.00	77	0.56	7.37
O	3.78	12	21.42	20	0.44	4.33	3.31	7	15.90	17	0.61	3.78
P	20.58	132	60.55	80	0.52	8.14	22.81	131	54.58	86	0.50	6.43
Q	8.11	48	32.87	59	0.58	5.90	3.37	29	17.46	74	0.63	3.91
R	8.53	46	79.31	20	0.40	5.00	7.05	25	67.57	7	0.28	3.11
S	2.21	22	13.84	69	0.54	3.89	7.03	31	23.48	61	0.55	3.51
T	21.57	137	72.11	85	0.52	11.93	22.58	134	74.44	66	0.48	8.60
Mean (range)	7.85 (1.62-21.57)	42.45 (0-137)	39.89 (0-79.31)	38.65 (0-164)	0.42 (0-0.59)	4.60 (1-11.93)	8.91 (1.68-22.81)	37.65 (0-134)	40.12 (0-100)	37.45 (0-86)	0.47 (0-0.63)	3.85 (1-8.6)